

**SULIT**



First Semester Examination  
Academic Session 2018/2019

December 2018/January 2019

**MST567 – Categorical Data Analysis**  
***(Analisis Data Berkategori)***

Duration : 3 hours  
*[Masa : 3 jam]*

Please check that this examination paper consists of EIGHT (8) pages of printed material before you begin the examination.

*[Sila pastikan bahawa kertas peperiksaan ini mengandungi LAPAN (8) muka surat yang bercetak sebelum anda memulakan peperiksaan ini.]*

**Instructions** : Answer **SIX (6)** questions.

**[Arahan** : Jawab **ENAM (6)** soalan.]

In the event of any discrepancies, the English version shall be used.

*[Sekiranya terdapat sebarang percanggahan pada soalan peperiksaan, versi Bahasa Inggeris hendaklah digunapakai].*

...2/-

**SULIT**

**Question 1**

- (a) Universiti Sains Malaysia (USM) use letter grades to assess student performance but produce a grade point average (GPA) based on numbers associated with the letter grades. For example, an B+ is associated with a value of 3.33 for the calculation of the GPA. What is the respective levels of measurements for GPA? Under what assumptions is this procedure correct?

[ 4 marks ]

- (b) Thirty percent of all people who are inoculated with the current vaccine that is used to prevent a disease contract the disease within a year. The developer of a new vaccine that is intended to prevent this disease wishes to test for significant evidence that the new vaccine is more effective. Suppose twenty people inoculated with the new vaccine are studied and only three people contract the disease within a year. Can the developer conclude that the new vaccine more effective?

[ 6 marks ]

**Soalan 1**

- (a) *Universiti Sains Malaysia (USM) menggunakan gred abjad untuk menilai prestasi pelajar tetapi menghasilkan purata titik gred (GPA) berdasarkan nilai yang berkaitan dengan gred abjad tersebut. Contohnya, B+ dikaitkan dengan nilai 3.33 untuk pengiraan GPA. Apakah tahap pengukuran GPA? Di bawah anggapan apakah prosedur ini betul?*

[ 4 markah ]

- (b) *Tiga puluh peratus daripada semua orang yang disuntik dengan vaksin semasa yang digunakan untuk mengelakkan penyakit mengidap penyakit dalam tempoh setahun. Pengeluar vaksin baru yang bertujuan untuk mencegah penyakit ini ingin menguji bahawa vaksin baru ini lebih berkesan secara bererti. Kira-kira dua puluh orang disuntik dengan vaksin baru dikaji dan hanya tiga orang yang mengidap penyakit dalam tempoh setahun. Bolehkah pembangun menyimpulkan bahawa vaksin baru lebih berkesan?*

[ 6 markah ]

**Question 2**

In a random sample of 996 individuals were classified according to their marital status and whether they considered life dull, routine, or exciting. The results are shown in Table 1.

- (i) Partition Table 1 into appropriate number of sub tables.
- (ii) For each sub tables, conduct an inferential analysis using odds ratio. Interpret the results.

Table 1

		Attitude about Life			
		Dull	Routine	Exciting	Total
Marital Status	Married	21	241	251	<b>513</b>
	Widowed	17	54	40	<b>111</b>
	Divorced	10	74	65	<b>149</b>
	Separated	6	11	8	<b>25</b>
	Never married	11	79	108	<b>198</b>
	<b>Total</b>	<b>65</b>	<b>459</b>	<b>472</b>	<b>996</b>

[ 20 marks ]

**Soalan 2**

Dalam suatu sampel rawak dengan 996 individu dikelaskan mengikut status perkahwinan mereka dan sama ada mereka menganggap kehidupan membosankan, rutin, atau menarik. Keputusan ditunjukkan dalam Jadual 1

- (i) Partisi Jadual 1 berdasarkan bilangan sub jadual yang sesuai.
- (ii) Untuk setiap sub jadual, jalankan analisis pentaabiran menggunakan nisbah odds. Terangkan keputusannya.

Jadual 1

		Sikap tentang Kehidupan			
		Bosan	Rutin	Menarik	Jumlah
Status Perkahwinan	Berkahwin	21	241	251	<b>513</b>
	Balu	17	54	40	<b>111</b>
	Berceraai	10	74	65	<b>149</b>
	Berpisah	6	11	8	<b>25</b>
	Tidak berkahwin	11	79	108	<b>198</b>
	<b>Jumlah</b>	<b>65</b>	<b>459</b>	<b>472</b>	<b>996</b>

[ 20 markah ]

**Question 3**

A diagnostic test is 92% sensitivity and 94% specificity. A test group is comprised of 850 people known to have the disease and 850 people known to be free of the disease.

- (i) How many of the known positives would actually test positive?
- (ii) How many of the known negatives would actually test negative?

...4/-

(iii) Calculate the Youden index and interpret the value.

(iv) Find the positive and negative predictive values.

[ 10 marks ]

**Soalan 3**

*Suatu ujian diagnostik adalah sensitiviti 92% dan kekhususan 94%. Suatu kumpulan ujian terdiri daripada 850 orang yang diketahui mempunyai penyakit dan 850 orang yang diketahui bebas daripada penyakit ini.*

*(i) Berapa banyak positif yang diketahui sebenarnya akan berkeputusan positif?*

*(ii) Berapa banyak negatif yang diketahui sebenarnya akan berkeputusan negatif?*

*(iii) Kira indeks Youden dan tafsir nilai tersebut.*

*(iv) Cari nilai ramalan positif dan negatif.*

[ 10 markah ]

**Question 4**

For the 24 space shuttle flights before the Challenger mission disaster in 1986, Table 2 shows two variables: Temp, the outside temperature (in degrees Fahrenheit) at the time of the launch, and BadRings, the number of O-rings that showed signs of thermal distress following the launch. An O-ring is a seal that separates the fuel supply from the combustible gases in the rocket's exhaust; if it fails to do so perfectly, it will show signs of thermal distress after the launch. In cold weather, O-rings are less resilient and may be more likely to fail.

Table 2

Temp	BadRings	Temp	BadRings	Temp	BadRings	Temp	BadRings
53	3	67	0	70	2	76	0
56	2	67	0	70	1	76	0
57	2	68	0	72	0	78	0
63	0	69	0	73	0	79	0
66	0	70	0	75	0	80	0
67	0	70	1	75	2	81	0

(i) Using logistic regression, model the way in which the probability of an O-ring failure depends on temperature.

...5/-

- 5 -

- (ii) Test the null hypothesis that O-ring failures are independent of temperature. Does this seem plausible given the data?
- (iii) Estimate the odds ratio for O-ring failure with a 10 degree decrease in the launch temperature.
- (iv) The 25th space shuttle launch, involving the space shuttle Challenger, took place on January 27, 1986. Seventy-three seconds into the flight, the fuel mixed with the rocket exhaust, resulting in an explosion which destroyed the shuttle and killed all seven astronauts on board. The launch temperature that day was 31 degrees. Based on data from the first 24 launches, estimate the probability of an O-ring failure on the Challenger flight.
- (v) Provide a confidence interval for the probability in part (iv).

[ 20 marks ]

**Soalan 4**

Bagi 24 pesawat ke ruang angkasa sebelum bencana misi Challenger pada tahun 1986, Jadual 2 menunjukkan dua pemboleh ubah: Temp, suhu luar (darjah Fahrenheit) pada masa pelancaran, dan BadRings, bilangan O-ring yang menunjukkan tanda-tanda gangguan terma selepas pelancaran. O-ring adalah penutup yang memisahkan bekalan bahan api daripada gas mudah terbakar dalam ekzos roket; jika ia gagal melakukannya dengan sempurna, ia akan menunjukkan tanda-tanda gangguan terma selepas pelancarannya. Dalam cuaca sejuk, O-ring kurang tahan lasak dan mungkin lebih cenderung gagal.

Jadual 2

Temp	BadRings	Temp	BadRings	Temp	BadRings	Temp	BadRings
53	3	67	0	70	2	76	0
56	2	67	0	70	1	76	0
57	2	68	0	72	0	78	0
63	0	69	0	73	0	79	0
66	0	70	0	75	0	80	0
67	0	70	1	75	2	81	0

- (i) Gunakan regresi logistik, model dengan cara di mana kebarangkalian kegagalan O-ring bergantung kepada suhu.
- (ii) Uji hipotesis nol bahawa kegagalan O-ring adalah tidak bergantung kepada suhu. Adakah ini wajar berdasarkan data yang diberikan?
- (iii) Anggarkan nisbah odds untuk kegagalan O-ring dengan pengurangan 10 darjah suhu pelancaran.

...6/-

- (iv) *Pelancaran pesawat ke ruang angkasa ke-25, yang melibatkan pesawat angkasa Challenger, berlaku pada 27 Januari 1986. Tujuh puluh tiga saat dalam penerbangan, bahan api bercampur dengan ekzos roket, mengakibatkan letupan yang menghancurkan pesawat dan membunuh semua tujuh angkasawan dalam pesawat. Suhu pelancaran hari itu ialah 31 darjah. Berdasarkan data dari 24 pelancaran pertama, anggarkan kebarangkalian kegagalan O-ring pada penerbangan Challenger.*
- (v) *Berikan selang keyakinan untuk kebarangkalian pada (iv).*

[ 20 markah ]

**Question 5**

Table 3 depicts data on respondents' gender, level of fundamentalism, and whether they favor or oppose the death penalty for murder. Fit and interpret the most parsimonious log-linear model to this data. Describe the process you went through to determine the best fitting, most parsimonious model. What types of associations exist? Demonstrate how the fitted values obtained from the model depict these associations.

Table 3

		Favor or oppose death penalty for murder		
Fundamentalism	Gender	Favor	Oppose	Total
Fundamentalist	Male	128	32	160
	Female	123	73	196
	Total	251	105	356
Moderate	Male	182	56	238
	Female	168	105	273
	Total	350	161	511
Liberal	Male	119	49	168
	Female	111	70	181
	Total	230	119	349

[ 20 marks ]

...7/-

**Soalan 5**

Jadual 3 menggambarkan data jantina responden, tahap fundamentalisme, dan sama ada mereka menyokong atau menentang hukuman mati untuk pembunuh. Suaikan dan tafsirkan model log-linear yang paling parsimoni kepada data ini. Terangkan proses yang anda jalankan untuk menentukan model yang paling sesuai, paling parsimoni. Apakah jenis hubungan yang ada? Tunjukkan bagaimana nilai suai yang diperoleh daripada model menggambarkan hubungan ini.

*Jadual 3*

		Menyokong atau menentang hukuman mati untuk pembunuh		
<i>Fundamentalisme</i>	<i>Jantina</i>	<i>Menyokong</i>	<i>Menentang</i>	<i>Jumlah</i>
<i>Fundamentalis</i>	<i>Lelaki</i>	128	32	160
	<i>Wanita</i>	123	73	196
	<i>Jumlah</i>	251	105	356
<i>Sederhana</i>	<i>Lelaki</i>	182	56	238
	<i>Wanita</i>	168	105	273
	<i>Jumlah</i>	350	161	511
<i>Liberal</i>	<i>Lelaki</i>	119	49	168
	<i>Wanita</i>	111	70	181
	<i>Jumlah</i>	230	119	349

[ 20 markah ]

**Question 6**

Based on Table 3 from Question 5, choose a research question that can be addressed with these data using a binary logistic regression analysis. Explain which of the variables is to be used as the outcome and which as predictors. Fit the binary logistic regression model to address your research question. Fully interpret the results of the analysis, including the process you used to select a final model and substantive interpretations of its parameter estimates. Show the SAS program used to analyze the model.

[ 20 marks ]

...8/-

**Soalan 6**

*Berdasarkan Jadual 3 daripada Soalan 5, pilih soalan penyelidikan yang boleh ditangani dengan data ini dengan menggunakan analisis regresi logistik binari. Jelaskan pembolehubah mana yang digunakan sebagai dapatan dan yang mana sebagai peramal. Suaikan model regresi logistik binari untuk menangani soalan penyelidikan anda. Terangkan sepenuhnya hasil analisis, termasuk proses yang anda gunakan untuk memilih model akhir dan tafsiran substantif anggaran parameteranya. Tunjukkan program SAS yang digunakan untuk menganalisis model.*

*[ 20 markah ]*

**- ooo00ooo -**